DOI: 10.1534/genetics.108.090068

Note

Patterns of Amino Acid Evolution in the *Drosophila ananassae* Chimeric Gene, *siren*, Parallel Those of Other *Adh*-Derived Chimeras

Hung-Jui Shih and Corbin D. Jones[1]

Department of Biology and Carolina Center for the Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599-3280

Manuscript received April 7, 2008
Accepted for publication July 22, 2008

ABSTRACT

siren1 and *siren2* are novel *alcohol dehydrogenase* (*Adh*)-derived chimeric genes in the *Drosophila bipectinata* complex. *D. ananassae*, however, harbors a single homolog of these genes. Like other *Adh*-derived chimeric genes, *siren* evolved adaptively shortly after it was formed. These changes likely shifted the catalytic activity of *siren*.

GENE duplication is a major source of new genes (Ohno 1970). Some duplications, such as fusion genes (FGs), involve radical reorganizations of gene structures, which result in immediate functional divergence. FGs occur when two previously separate and independent genes are fused to form a single chimeric ORF. Recently, Nozawa *et al.* (2005) discovered a pair of new *Adh*-derived FGs in the *Drosophila bipectinata* complex—*siren1* and *siren2*. They showed that the ancestor to these genes was likely formed by the retrotransposition of a fragment of *CG11779* in front of a duplicated *Adh* gene. Nozawa *et al.* (2005), however, did not identify *siren* in *D. ananassae*, an outgroup of the *D. bipectinata* complex (Kopp and True 2002; Kopp and Barmina 2005; Kopp 2006; Schug *et al.* 2007).

An *Adh* duplication in *D. ananassae* was bioinformatically identified in a preliminary draft of the *D. ananassae* genome (Jones and Begun 2005). Analysis of regions flanking the duplicated *Adh* show strong similarity to the 5′-ends of *siren1* and *siren2*. We experimentally confirmed that *D. ananassae siren* was single copy (Figure 1).

D. ananassae siren is actively transcribed (see supplemental Methods). Like the 5′-UTR of *siren1* in *bipectinata* (Nozawa *et al.* 2005), the *D. ananassae siren* 5′-UTR extends at least 264 bases 5′ of the *siren* start codon. *D. ananassae siren* is highly expressed in males (Figure 1). In contrast to *siren1* and *siren2* of the *D. bipectinata* complex, *D. ananassae siren* is also expressed in females, although at a significantly lower level than in males (Figure 1).

The protein sequence of SIREN is substantially different from that of its ancestors. The ADH-derived region of SIREN is more diverged from *D. melanogaster* ADH than any published Drosophila ADH protein (58 of 256 amino acids) suggesting that SIREN does not retain ADH-like activity. Prior allozyme studies of ADH in *D. ananassae* provided no evidence of a protein other than ADH that oxidized ethanol (Johnson *et al.* 1966; Stone *et al.* 1968). Using two independent ADH antibodies (ADH-Ab), Western blots revealed a single band of the size expected for ADH in *D. melanogaster* and *D. ananassae*, but no strong secondary band of the size expected for SIREN. This lack of cross-reactivity with ADH-Ab is typical for FGs and reflects their radical reorganization (Zhang *et al.* 2004; Jones *et al.* 2005).

To better understand the evolutionary mechanisms affecting *siren*, we surveyed the DNA polymorphism present in the second and third exons of *siren* (supplemental Figure 1). Other than the first 168 bases, this region is derived from *Adh*. We also surveyed sequence variation in the equivalent region of *D. ananassae Adh* (Table 1). *D. ananassae siren* is not highly divergent from *siren1* and *siren2* (supplemental Figure 1). The comparison of *D. ananassae siren* and *D. ananassae Adh* shows that patterns of sequence polymorphism are qualitatively similar, but that divergence in the catalytic domain is more constrained (supplemental Figure 1C). We found no evidence for recent adaptive evolution between *siren* and *siren1* or *siren2* [McDonald–Kreitman test (McDonald and Kreitman 1991); *siren1* *G*-test, *G*-value = 0.760, $P = 0.383$; *siren2* *G*-test, *G*-value = 1.097, $P = 0.295$].

Sequence data from this article have been deposited with GenBank Data Libraries under accessions nos. EU877936–EU877945 and EU880420–EU880429.

[1]*Corresponding author:* Department of Biology, Carolina Center for Genome Sciences, Campus Box 3280, Coker Hall, UNC, Chapel Hill, NC 27599-3280. E-mail: cdjones@email.unc.edu

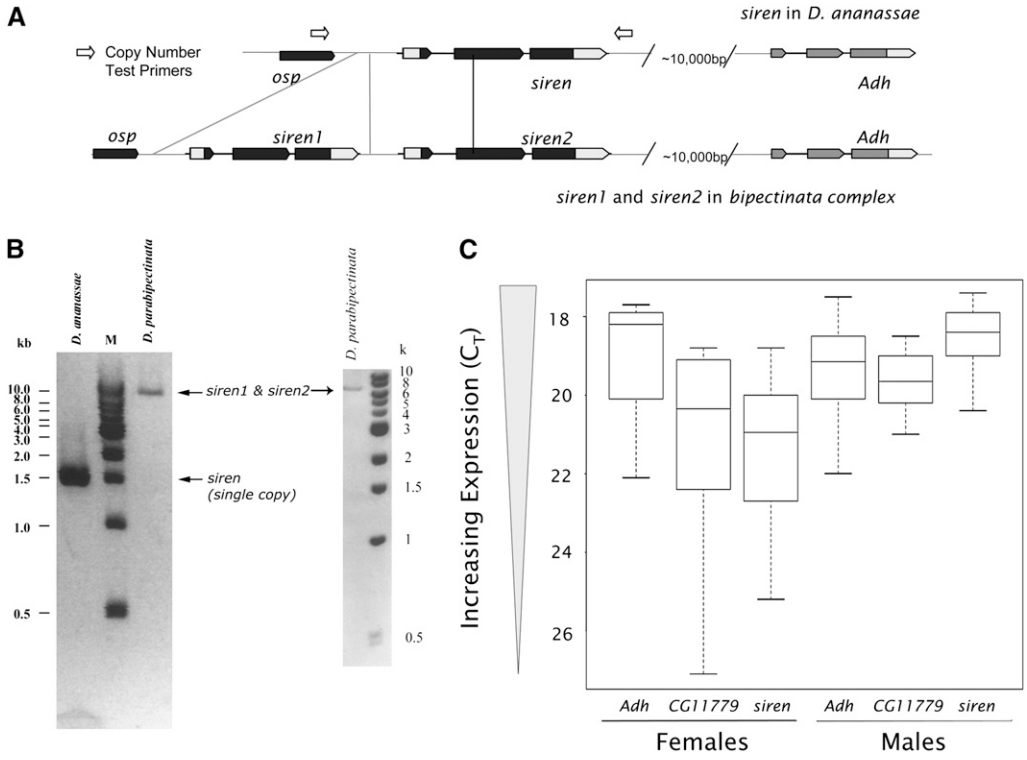

FIGURE 1.—(A) Comparison of the *siren* region in *D. ananassae* to that of *siren1* and *siren2* region in the *bipectinata* complex, including locations of primers used to confirm the number of copies of *siren*. The 5′ primer is located within the coding region of the gene *outspread* (*osp*); the 3′ primer is located in a region of unique sequence 3′ from the end of the transcribed region of *siren2*. (B) PCR amplicons using identical primers ("copy number test" in A). These primers amplify bands of different sizes in different species. (Left) A *D. ananassae* band much smaller than the *D. parabipectinata* band. The *D. ananassae* amplicon is ~1500 nt, which is consistent with our expectation from the genome analysis. (Right) The same primers amplify a much larger fragment in *D. parabictectinata*, which is the size expected for two copies of *siren*. Thus, *siren* is present in only one copy in *D. ananassae*. We confirmed this result with a combination of DNA sequencing of the PCR amplicon and Southern blot analysis (supplemental Methods). (C) Relative to *Adh* and CG11779, *siren* is more expressed in males. C_T (cycle times) represents the number of cycles required before the fluorescence of the amplicons crosses the baseline fluorescence threshold. C_T on the *y*-axis is plotted in reverse order as lower C_T implies increased transcript abundance in the initial sample. Box-and-whisker plot shows median and quartiles of C_T values for a population of 10 different *D. ananassae* lines. For each sex of each line, the same cDNA sample was used for all replicates in all genes. These data were analyzed with an ANOVA (model: CT = μ + GENE + SEX + LINE + GENE:SEX + GENE:LINE + LINE:SEX + GENE:SEX:LINE + ε). In addition to the main effects, there is a significant GENE:SEX interaction (P= 0.0099), which after post-hoc Tukey's significant difference analysis results from the difference between *siren* expression in females compared to males (P <0.0001).

Following JONES and BEGUN (2005), we used a contingency table approach to compare paralogous loci within species to detect dramatic changes in the substitution rate at synonymous and nonsynonymous sites. Our analysis suggests that the *Adh* region of *siren* has fixed a disproportionate number of nonsynonymous changes (bases surveyed, 771; synonymous substitutions, 29; nonsynonymous substitutions, 59; synonymous polymorphisms, 8; nonsynonymous polymorphisms, 4; *G*-test, *G*-value = 4.96, $P = 0.026$). Given that extant *sirens* are evolutionarily constrained, the contingency table analysis suggests that *siren* rapidly diverged from its *Adh* ancestor and that much of the amino acid divergence may have arisen prior to the speciation of *D. ananassae* and the ancestor of the *bipectinata* group. This scenario would lead to an elevated rate of amino acid substitution (d_N) relative to the background rate of synonymous substitution (d_S) in the *Adh* region of *siren* compared to the same regions in *Adh* (YANG 1997; YANG *et al.* 2000; JONES *et al.* 2005; see supplemental Methods). Analysis of d_N/d_S shows that the initial bout of adaptive evolution occurred after *siren* split from *Adh* (supplemental Figure 2). Afterward, evolution at *siren* slowed. None of the other *siren* lineages evinces adaptive evolution, although the rate of amino acid divergence remains high along the branch leading to *D. ananassae* and in the branch prior to the speciation of the *bipectinata* complex (supplemental Figure 2; supplemental Table 1).

TABLE 1

Nucleotide polymorphism statistics for *siren* and *Adh* in *D. ananassae*

	siren	*Adh*
No. of alleles	10	10
Bases surveyed	1027	859
Total sites (excluding gaps)	888	771
No. polymorphic (segregating) sites, *S*	16	17
No. haplotypes, N_{hap}	5	10
Haplotype diversity	0.66	1.00
Polymorphism		
Nucleotide diversity, *P*	0.00623	0.0077
$\theta_{per\ site}$	0.00637	0.00825
Mean no. of differences among *D. ananassae* alleles	5.5	6

TABLE 2

Amino acid substitutions in *siren* often occur at the same positions as other *Adh*-derived fusion genes

	% shared amino acid changes	*P*-value *vs.* expectation[b]
Comparison of *D. ananassae siren*[a]		
vs. jingwei	63	0.0042
vs. Adh-Finnegan	71	<0.0001
vs. Adh-Twain	71	0.0002
Comparison of *siren* ancestor		
vs. jingwei	36	0.066
vs. Adh-Finnegan	50	0.0001
vs. Adh-Twain	50	0.0024

[a] We ignored the two-amino-acid indel difference among these species. A total of 254 amino acid positions were analyzed.

[b] Fisher's exact test expectation is 22 and 14%, respectively.

Prior work suggests that independently derived *Adh* FGs show parallel patterns of amino acid evolution and that this parallelism is *not* the product of neutral evolutionary processes (Jones and Begun 2005). Patterns of early adaptive amino acid evolution at *siren* parallel those observed in *Adh*-derived FGs, *jingwei*, *Adh-Finnegan*, and *Adh-Twain* (Jones and Begun 2005). In all three cases, *siren* and the fusion genes evolved at many of the same positions (Table 2). These parallel, adaptively evolving sites in FGs also tend to be highly functionally constrained sites in *Adh* (Jones and Begun 2005). Eleven of the potentially adaptive substitutions in *D. ananassae siren* (5 prior to speciation of *D. ananassae* and the *bipectinata* complex; 6 after this speciation event) occur at sites that never evolve in Drosophila *Adh*.

Directional selection clearly drove early divergence at *siren*, often at the same sites as other FGs. This pattern may have been driven by a need to evolve away from the ancestral protein's activity and expression (Jones and Begun 2005). When formed, these four *Adh* FGs recruited new 5′ regulatory sequences, which resulted in novel expression patterns. If the fused *Adh* region initially retained some *Adh*-like activity, the FG might express this activity at inappropriate times or in the wrong tissues. This is potentially deleterious for the fly (*e.g.*, Parsch *et al.* 2000), even if the FG is largely beneficial. Mutations that ameliorate this pleiotropic effect by removing the ancestral *Adh*-like activity would be advantageous. Our data from *siren* are consistent with this hypothesis: the expression pattern of *siren* differs from *Adh*, functionally important sites in *Adh* were altered in *siren* shortly after it formed, and natural selection likely drove early divergence in *siren*. Thus, selection to reduce deleterious pleiotropy likely drives early evolution of FGs.

We thank Dave Begun, Josie Reinhardt, Cass Jabara, Alisha Holloway, Montse Aguadé, and two reviewers for thoughts on the manuscript. We thank Patricia Sheridan and Melinda Beck and Clinical Nutrition Research Unit for help with the quantitative PCR. We thank C. "Tip" Benyajati (and Uncle Billy) for the *Adh* antibody. Stocks were provided by the Drosophila Species Stock Center. This work was supported by the University of North Carolina, the Carolina Center for Genome Sciences, and a grant from National Institutes of Health (DK56350) to the University of North Carolina Clinical Nutrition Research Unit.

LITERATURE CITED

Johnson, F. M., C. G. Kanapi, R. H. Richardson, M. R. Wheeler and W. S. Stone, 1966 An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. Proc. Natl. Acad. Sci. USA **56:** 119–125.

Jones, C. D., and D. J. Begun, 2005 Parallel evolution of chimeric fusion genes. Proc. Natl. Acad. Sci. USA **102:** 11373–11378.

Jones, C. D., A. W. Custer and D. J. Begun, 2005 Origin and evolution of a chimeric fusion gene in *Drosophila subobscura*, *D. madeirensis* and *D. guanche*. Genetics **170:** 207–219.

Kopp, A., 2006 Basal relationships in the *Drosophila melanogaster* species group. Mol. Phylogenet. Evol. **39:** 787–798.

Kopp, A., and O. Barmina, 2005 Evolutionary history of the *Drosophila bipectinata* species complex. Genet. Res. **85:** 23–46.

Kopp, A., and J. R. True, 2002 Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst. Biol. **51:** 786–805.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Nozawa, M., T. Aotsuka and K. Tamura, 2005 A novel chimeric gene, *siren*, with retroposed promoter sequence in the *Drosophila bipectinata* complex. Genetics **171:** 1719–1727.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Parsch, J., J. A. Russell, I. Beerman, D. L. Hartl and W. Stephan, 2000 Deletion of a conserved regulatory element in the Drosophila *Adh* gene leads to increased alcohol dehydrogenase activity but also delays development. Genetics **156:** 219–227.

Schug, M. D., S. G. Smith, A. Tozier-Pearce and S. F. McEvey, 2007 The genetic structure of *Drosophila ananassae* populations from Asia, Australia and Samoa. Genetics **175:** 1429–1440.

Stone, W. S., M. R. Wheeler, F. M. Johnson and K. I. Kojima, 1968 Genetic variation in natural island populations of members of the *Drosophila nasuta* and *Drosophila ananassae* subgroups. Proc. Natl. Acad. Sci. USA **59:** 102–109.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Zhang, J., A. M. Dean, F. Brunet and M. Long, 2004 Evolving protein functional diversity in new genes of Drosophila. Proc. Natl. Acad. Sci. USA **101:** 16246–16250.

Communicating editor: M. Aguadé